



OPEN

DATA DESCRIPTOR

A 100-member ensemble simulations of global historical (1951–2010) wave heights

Mercè Casas-Prat¹✉, Xiaolan L. Wang¹, Nobuhito Mori², Yang Feng¹, Rodney Chan¹ & Tomoya Shimura²

The d4PDF-WaveHs dataset represents the first single model initial-condition large ensemble of historical significant ocean wave height (H_s) at a global scale. It was produced using an advanced statistical model with predictors derived from Japan's d4PDF ensemble of historical simulations of sea level pressure. d4PDF-WaveHs provides 100 realizations of H_s for the period 1951–2010 (hence 6,000 years of data) on a $1^\circ \times 1^\circ$ lat.-long. grid. Technical comparison of model skill against modern reanalysis and other historical wave datasets was undertaken at global and regional scales. d4PDF-WaveHs provides unique data to understand better the poorly known role of internal climate variability in ocean wave climate, which can be used to estimate better trend signals. It also provides a better sampling of extreme events. Overall, this is crucial to properly assess wave-driven impacts, such as extreme sea levels on low-lying populated coastal areas. This dataset may be of interest to a variety of researchers, engineers and stakeholders in the fields of climate science, oceanography, coastal management, offshore engineering, and energy resource development.

Background & Summary

Ocean wind-waves, hereafter called waves, are an important element of the climate system, modulating the interactions between the atmosphere and the oceans¹. They are also a key environmental variable for coastal and offshore engineering², and affect many coastal dynamics processes³, navigation planning⁴, and are a potential source of renewable energy⁵. Over 300 million people live in low-lying coastal areas⁶, and detailed knowledge of wave climate is essential to address the environmental and societal wave-driven impacts properly.

The IPCC (2013)⁷ highlighted a low knowledge confidence of wave climatology in comparison with many other climate variables. This relates to the fact that most climate models provide no information about waves; therefore, the availability of wave simulations is relatively limited. To fill in this gap, a growing number of studies have been developed over the last decade, producing several global and regional wave datasets. Most of these efforts were consolidated with COWCLIP2.0⁸, the first coherent, community-driven multi-method ensemble of historical and future global wave simulations, which included dominant sources of uncertainty, namely forcing uncertainty, and wave and climate model uncertainty. However, the internal climate variability was not properly sampled as most combinations of forcing and climate/wave models considered just one realization of the climate system. Studies based on a single (or reduced sample of) realizations of the climate system might underestimate extreme events or confound trends with internal climate variability^{9,10}.

More recently, a global ensemble of ocean wave climate statistics from contemporary wave reanalysis and hindcasts¹¹ highlighted the discrepancies among modern wave products of historical data. However, this database cannot provide insight into the role of the internal climate variability either due to the relatively short time period considered (1980–2014). Also, wave reanalysis/hindcasts are constructed to replicate the observed climate and, therefore, they all correspond to the same realization of the climate system.

The internal climate variability can be investigated with a Single Model Initial-condition Large Ensemble (SMILE), which is a set of simulations starting from different initial conditions but produced with a single climate model and identical external forcing¹². Over the last decade, SMILEs have been increasingly generated and used in climate science as they represent very valuable data to study not just internal climate variability but also extremes^{13,14}. The number of ensemble members required to obtain robust estimates depends on targets or

¹Climate Research Division, Science and Technology Branch, Environment and Climate Change Canada, Toronto, Ontario, Canada. ²Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan. ✉e-mail: merce.casasprat@ec.gc.ca

Statistics ID	Indicator name	Time-frame resolution	Units
Hs_avg	Mean significant wave height	Annual (1), Seasonal (4) and Monthly (12)	m
Hs_p10	10th Percentile significant wave height	Annual (1), Seasonal (4) and Monthly (12)	m
Hs_p50	50th Percentile significant wave height	Annual (1), Seasonal (4) and Monthly (12)	m
Hs_p90	90th Percentile significant wave height	Annual (1), Seasonal (4) and Monthly (12)	m
Hs_p95	95th Percentile significant wave height	Annual (1), Seasonal (4) and Monthly (12)	m
Hs_p99	99th Percentile significant wave height	Annual (1), Seasonal (4) and Monthly (12)	m
Hs_max	Maximum significant wave height	Annual (1), Seasonal (4) and Monthly (12)	m

Table 1. d4PDF-WaveHs set of H_s statistics (via *getStat.f*).

Statistics ID	Indicator name	Definition	Units
HsRo	Rough wave days	Annual count of days when daily max $H_s > 2.5$ m	days
HsHi	High wave days	Annual count of days when daily max $H_s > 6$ m	days
fHsRo	Frequency of rough wave days	Annual percentage of days when daily max $H_s > 2.5$ m	%
fHsHi	Frequency of high wave days	Annual percentage of days when daily max $H_s > 6$ m	%
fHs10p	Frequency of top decile wave days	Annual percentage of days when daily max $H_s > 10$ th percentile of daily max H_s in base period*	%
fHs90p	Frequency of top decile wave days	Annual percentage of days when daily max $H_s < 90$ th percentile of daily max H_s in base period*	%
HHsDI	Top decile wave spell duration indicator	Annual count of days with at least 2 consecutive days when daily max $H_s > 90$ th percentile of daily max H_s in the base period*	days

Table 2. d4PDF-WaveHs set of H_s statistics (via *getHsEx.f*). *Base period is 1980–2010.

temporal and spatial averaging scale¹⁵. However, most studies conclude that single realizations are insufficient to assess climate statistics and that large samples (size of >20 – 30 members) are required¹⁰. However, as it happens for other ensembles produced by climate models, existing SMILEs do not provide information about waves.

Here, we present and describe d4PDF-WaveHs, the first SMILE-based ensemble of global significant wave height (H_s) simulations. H_s is a well-defined and standardized statistic to describe the characteristic wave height of the sea state, which is defined as the average height of the highest one-third of waves, and it is largely used in coastal, naval, and offshore engineering. d4PDF-WaveHs was produced with an advanced statistical model^{16,17} and using d4PDF's historical simulations of sea level pressure (SLP) developed by the Japan Meteorological Research Institute with the MRI-AGCM atmospheric global climate model¹⁸. This dataset is archived in Network Common Data Form (NetCDF) with CF (Climate & Forecasts) compliant metadata, and contains 100 realizations of global H_s simulations over the period 1951–2010 on a 1° spatial grid resolution. The dataset provides a variety of standard H_s global statistics at monthly, seasonal, and annual time scale, as well as a set of extreme H_s indices designed by the Expert Team on Climate Change Detection (ETCCDI), using a standardized framework as in COWCLIP2.0 (see Tables 1, 2).

The d4PDF-WaveHs dataset provides valuable data that can be used to advance understanding of the poorly known role of the internal wave climate variability. This can result in a more robust assessment of H_s trends and low-frequency extremes. For instance, the annual set of wave statistics from the d4PDF-WaveHs ensemble was recently used to quantify the role of the internal climate variability (in comparison to other uncertainty factors) in the assessment of the annual mean and maximum H_s trends¹⁹. d4PDF-WaveHs can also contribute to improve broad-scale coastal hazard and vulnerability assessments. This extensive wave information can now be widely used by different stakeholders, engineers, and research communities, such as those focusing on natural hazards, coastal management, port, and offshore engineering, energy resource development and ship navigation.

Methods

This section describes the methodology used to generate the original sub-daily data, and the post-processed statistics and indices, which follows a standardized framework.

Generation of original sub-daily data. The modelling approach of Wang *et al.*^{16,17} was used to produce d4PDF-WaveHs. The main aspects of this advanced statistical model are summarized below but more details can be found in the reference papers^{16,17} and in the derived study that assesses trend uncertainty¹⁹. For each grid point of a global $1^\circ \times 1^\circ$ lat.-long. grid, a multivariate regression model with lagged dependent variable was developed. This regression model consists of expressing 6-hourly H_s as a function of 6-hourly anomalies (relative to the 1981–2000 mean) of SLP and of squared SLP gradients at each grid point, as well as the 30 leading principal components (PCs) of the SLP and of spatial SLP gradients fields over a large area of influence. 13 modelling regions were used (see Figure S1), which represents a slight modification of the original approach¹⁶, which used 11 modelling regions (ETNP and WTNP were a single region, as well as ETSP and WTSP). For each of these 13 regions a different set of PCs is considered. The consideration of local data accounts for local geostrophic wind energy, which drives local wind-sea states, while the set of PCs describe the large scale patterns of atmospheric

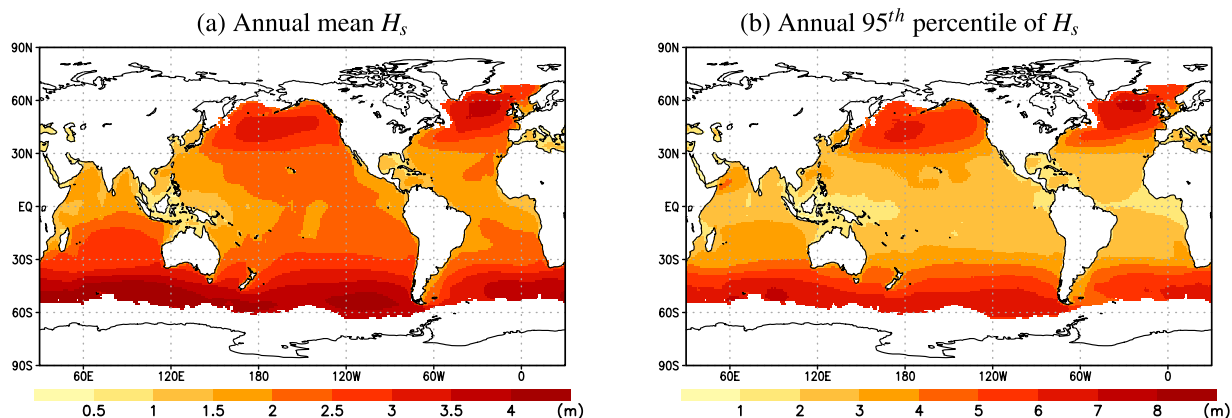


Fig. 1 Ensemble average of the annual mean H_s and 95th percentile of H_s climatological mean (m) as obtained from d4PDF-WaveHs for 1951–2010.

circulation that affect remotely generated (swell) waves arriving at a target grid point. In total, for each wave grid point, the model uses a pool of 62 potential SLP-based predictors and employs the F test with the equivalent sample size²⁰ to determine which and how these potential predictors are retained. Additionally, a Box-Cox power transformation²¹ is applied to both H_s and the predictors to minimize their departure from a normal distribution.

As in Wang *et al.*¹⁷, the European Center for Medium-Range Weather Forecasts (ECMWF) Reanalysis Interim (ERA-Interim) (ERA-Interim)²² was used to calibrate the model and bias-correct the predictors for each of the four seasons¹⁷. Particularly, the correction of predictors accounts for adjusting SLP fields to have the same climatological mean and standard deviation as ERA-Interim SLP data. Additionally, any simulated H_s values that exceed twice the largest H_s from ERA-Interim for a given season are excluded (set to missing). This cap is needed as very rarely (0.05% in all simulated H_s data) the Box-Cox transformation of the SLP gradients leads to an overgrowth of sharp SLP gradients that causes unrealistic H_s values. We used ERA-Interim as calibration dataset in order to be consistent with previous studies and, in particular, the ECCO(s) sub-ensemble of the COWCLIP2.0 dataset⁸. This enables a better comparison between the role of the internal climate variability and model uncertainty (e.g. for trend assessment¹⁹). However, considering the higher resolution and methodological advances implemented in the newer ERA5 dataset²³, we acknowledge that future studies should consider the re-calibration of the statistical model using ERA5.

To generate d4PDF-WaveHs, the aforementioned SLP-based predictors were obtained from the historical simulations of Japan's d4PDF large ensemble (which contains simulations of historical and future climates)¹⁸. d4PDF was obtained with the 60-km resolution MRI-AGCM model²⁴ developed by the Japan Meteorological Research Institute. The historical d4PDF SMILE-type ensemble used in this study was generated by perturbations of the historical sea surface temperature, sea ice concentration, and sea ice thickness in relation to the observed errors, while using the same forcing and global mean concentration of greenhouse gases. The d4PDF dataset (available at <https://diasjp.net/en/service/d4pdf-data-download>) satisfactorily simulates the past climate in terms of climatology, natural variations, and extreme events; and has been used in more than 70 papers¹⁵.

Computation of statistics and indices. To calculate the H_s statistics and indices for each ensemble member of the 6-hourly H_s simulations, we used a standardized framework similar to the protocol used for the COWCLIP2.0 dataset. The original sub-daily H_s data was post-processed with the COWCLIP Fortran code *getStat.f* and *getHsEx.f*²⁵ (after a slight modification to allow for missing data). We thus obtained a standard set of wave statistics (at annual, seasonal, and monthly time-frame resolutions). With *getStat.f*, we calculated seven wave H_s statistics: mean, 10th, 50th, 90th, 95th, 99th percentiles, and maximum (see Table 1). The seasonal statistics were computed on default seasons defined as DJF (December–January–February), MAM (March–April–May), JJA (June–July–August), and SON (September–October–November). Note that the d4PDF-WaveHs ice-covered areas were kept the same as those corresponding to ERA-Interim. As an example, Figure 1 shows the 1951–2010 climatological mean of the annual mean and 95th percentile of H_s , as derived from the d4PDF-WaveHs ensemble.

The *getHsEx.f* was used to calculate an ETCCDI set of extreme annual H_s indices, using the baseline period over 1980–2010 to compute the relative statistics. The output netCDF contains the following seven extreme statistics calculated annually: rough wave days, high wave days, frequency of high wave days, frequency of top (low) decile wave days, frequency of top (high) decile wave days, top decile wave spell duration indicator (see Table 2).

The data obtained with the Fortran code described above was post-processed with standard NetCDF operators (NCOs) for file manipulation, such as the “ncatted” command, to include all relevant metadata, including variables’ attributes: “long_name” and “units”, as well as several global attributes about the project, modelling centres, forcing and experiments configuration.

Data Records

The full archived dataset²⁶ comprising the statistics and indices described above (consult Methods section) can be accessed through the Open Data portal of the Government of Canada (<https://open.canada.ca/>), at DOI: <https://doi.org/10.18164/d68361d0-8141-48b9-a25e-a9bc98d71438>.

The data set in total comprises 400 files, with a total volume of 184 GB. We used a consistent directory structure and file naming with the following Data Reference Syntax (DRS):

- Directories: *d4PDF-WaveHs/historical/<ensemble_member>/<version>/<frequency>*
- Filenames: *Hs_glob_d4PDF-WaveHs_historical_<ensemble_member>_<frequency>_1951-2010.nc*
- where *<ensemble_member>* is of the form “rN”, being N the corresponding ensemble member that goes from 1 to 100. *<version>* is given in the form “vYYYYMM” (year/month) and *<frequency>* can be “ann”, “seas” or “mon”.

Recommended global attributes were defined and included accounting for the Attribute Convention for Dataset Discovery (ACDD) standards compliance. Note that although we followed many of the CF convention guidelines, the files are not strictly CF-compliant in time dimension - which uses units “years since” and “months since” the reference date, but this is consistent with the COWCLIP2.0 dataset⁸.

Technical Validation

The statistical modelling approach used to generate the sub-daily data in this study has been used and validated in several previous studies at global and regional scales, and to assess trends, projected changes, and variability^{16,17,27}. It was also used to generate the ECCC(s) sub-ensemble of the COWCLIP2.0 dataset⁸, which has led to improved understanding of the historical and future wave climates^{28,29}. Moreover, the d4PDF dataset used to compute the predictors has been extensively examined and assessed for model skill against satellite observations and reanalysis datasets in several studies¹⁵.

Additionally, the resulting model-skill of the presented d4PDF-WaveHs dataset was compared against similar global wave simulations and modern reanalysis. In particular, we considered the CMIP5-driven historical simulations of COWCLIP2.0 (CMIP5-COWCLIP)⁸, with special emphasis on the ECCC(s) sub-ensemble (CMIP5-ECCC(s)), and the global ensemble of ocean wave climate statistics from contemporary wave reanalysis and hindcasts¹¹. The latter includes 2 wave reanalysis (ERA1²², ERA5²³) and 12 wave hindcasts that are driven by 6 wind reanalysis products (for more details refer to Morim *et al.*¹¹):

- NCEP/NCAR-driven products: NCEPNCAR-IHC-GOW1.0³⁰
- CFSR-driven products: CFSR-CSIRO-G1D³¹, CFSR-CSIRO-CAWCR³², CFSR-IHC-GOW2.0³³, CFSR-JRC³⁴, CFSR-IFREMER³⁵
- ECMWF ERAI-driven products: ERAI-JRC³⁴, ERAI-NOC³⁶
- ECMWF ERA5-driven products: ERA5H^{37,38}
- JMA JRA-55-driven products: JR55-KU-ST2^{39,40}, JRA55-KU-ST4^{39,40}
- MERRA2-driven products: MERRA2-IORAS⁴¹

Additionally, we compared the first ensemble member of d4PDF-WaveHs with the corresponding H_s obtained with the traditional dynamical modelling approach. Specifically, the WAVEWATCH III version 5 with a spatial resolution of 0.5625°⁴² was run with the surface winds of the first d4PDF member. All datasets were compared altogether for the common period between 1980 and 2004.

The model performance was assessed in terms of the climatology distribution rather than the replication of particular weather events because climate simulations are not in phase with observations. In particular, the normalized version of Taylor diagram⁴³ was used for technical validation, which provides information about the spatial correlation, normalized standard deviation and normalized centred-root-mean-square difference of the H_s statistic of the model in question relative to the corresponding value of the reference dataset (the statistics were normalized, and non-dimensionalized, dividing by the standard deviation of the reference dataset). Since Taylor diagrams do not provide bias information, the technical validation also included comparison of the root-mean-square difference vs. the absolute difference (also relative to the reference dataset). The most recent reanalysis, ERA5, was used as reference dataset for both the Taylor diagrams and bias plots. Note this choice was made for comparison purposes only with the main goal to show the performance of the d4PDF-WaveHs in comparison to the variability among modern wave data products.

Figures 2, 3 show the performance of the annual mean and the annual 95th percentile of H_s climatologies at global scale. Additional results at regional scale are provided in the Supplementary Material (Figures S2–S21). For the ensemble of historical simulations, both the ensemble members and the ensemble average are illustrated in lighter and darker shades, respectively, of the same colour. Also, the wave hindcasts/reanalysis are grouped (in colour) in relation to the driving wind reanalysis. As expected, the performance of the global annual mean H_s obtained by d4PDF-WaveHs ensemble is similar to ECCC(s) and ERAI (Fig. 2a), as the latter was used in the calibration and predictors’ adjustment step of the statistical methodology used to simulate d4PDF-WaveHs and ECCC(s). These performances are also close to ERA5 with a relatively low RMSE and bias in comparison to the other wave products used for comparison. The largest RMSE is obtained for CFSR-CSIRO-G1D but the largest biases are obtained for a few CMIP5-COWCLIP members (Fig. 3a). For the annual 95th percentile of H_s , d4PDF-WaveHs tends to underestimate the corresponding ERA5 values but the overall performance is good in comparison to the overall uncertainty of all wave products (see Figs. 2b, 3b). Also, this negative bias is reduced for mid to high latitudes.

As expected, the internal climate variability (as described here by the d4PDF-WaveHs inter-member variability) of the climatological mean H_s is very small. A slightly larger spread is obtained for the 95th percentile, which further increases at regional scales (Figs. S3–S7). This inter-member variability is expected to further increase

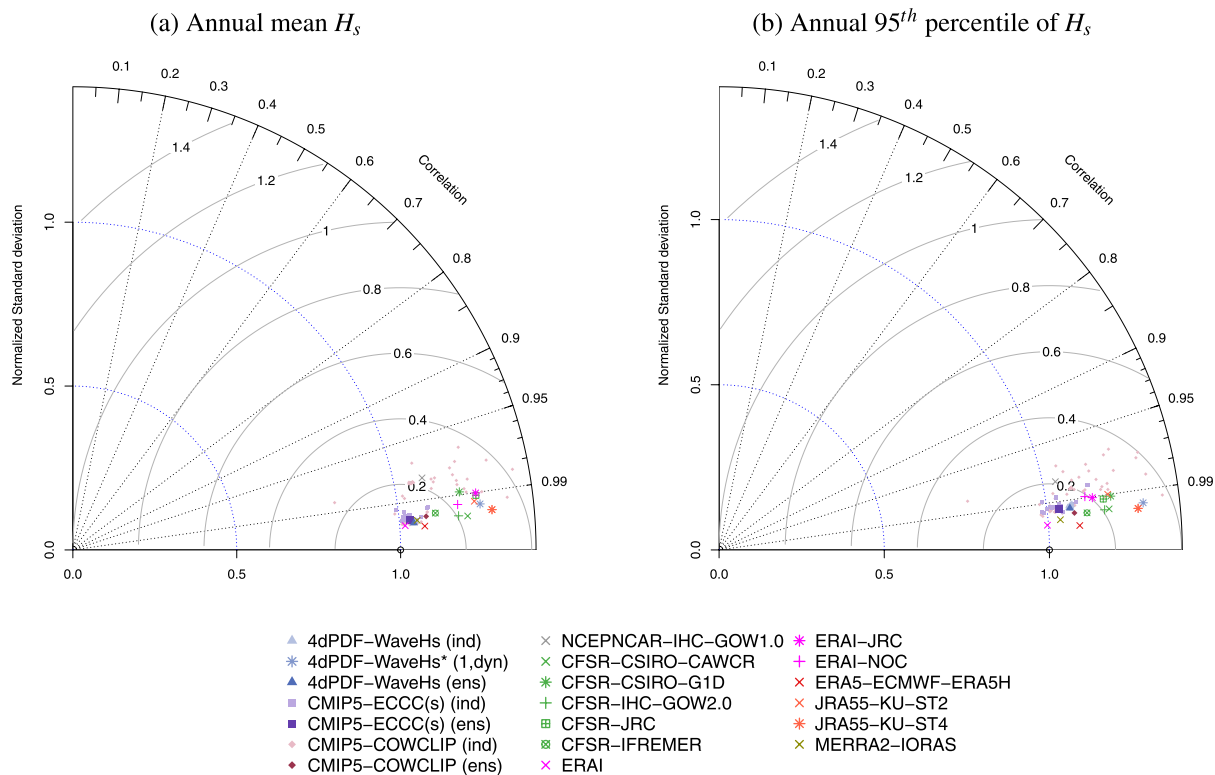


Fig. 2 Normalized Taylor Diagram for the annual mean H_s (a) and the annual 95th percentile of H_s (b) mean climatologies of the indicated wave products using ERA5 as reference. Normalized standard deviation is on the radial axis, correlation coefficient is on the angular axis and the gray lines indicate the normalized centred-root-mean-square difference.

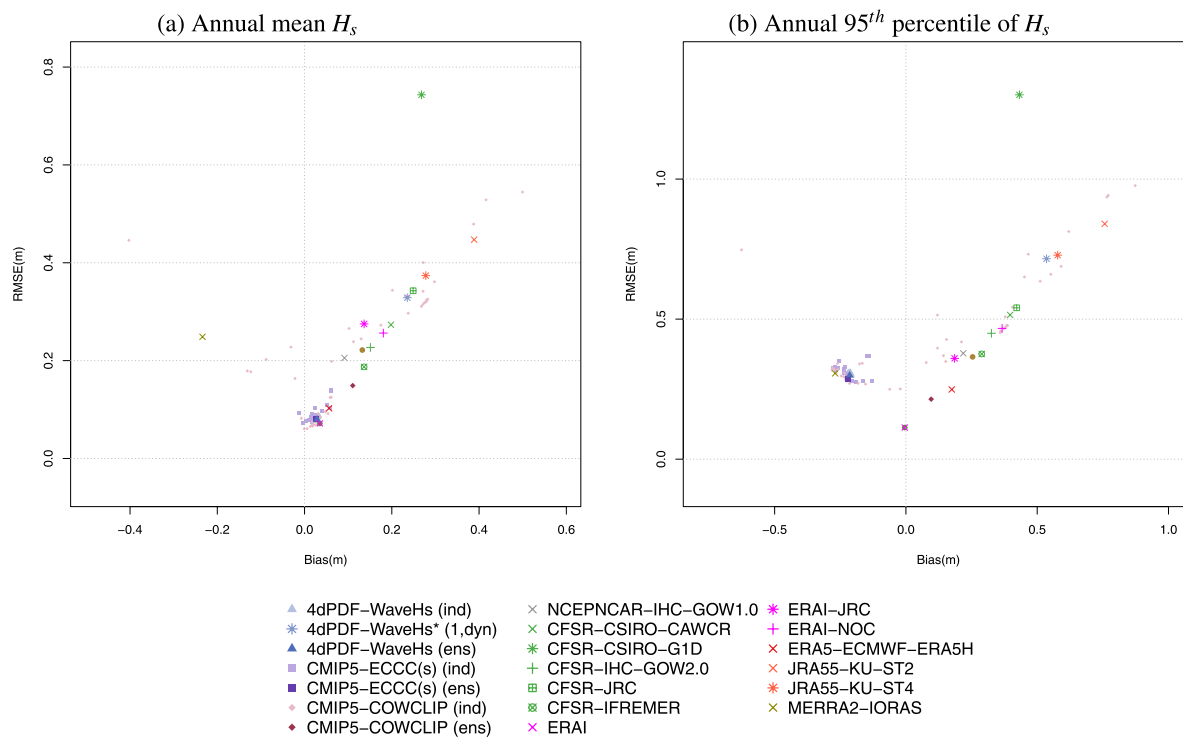


Fig. 3 RMSE (m) vs. Bias (m) for the annual mean H_s (a) and the annual 95th percentile of H_s (b) mean climatologies of the indicated wave products using ERA5 as reference.

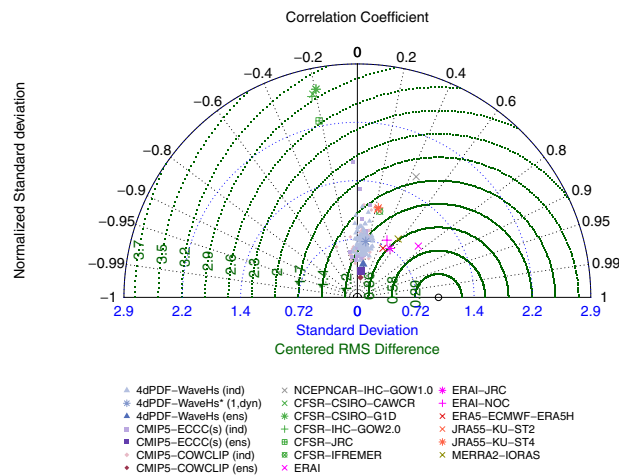


Fig. 4 Normalized Taylor Diagram for the annual mean H_s trend of the indicated wave products using ERA5 as reference. Normalized standard deviation is on the radial axis, correlation coefficient is on the angular axis and the dotted green lines indicate the normalized centred-root-mean-square difference. Note that the second quadrant on the left indicates negative correlations.

for lower frequency events. This is because the assessment of extremes is more sensitive to the internal climate variability than the average climate. The internal climate variability also plays an important role in the trend analysis. Figure 4 shows how the spread of trends of the annual mean H_s , corresponding to the d4PDF-WaveHs members is comparable to the variability of the corresponding values for the COWCLIP dataset, which includes the uncertainty derived from climate models and wave modelling approaches. Another interesting result from Figure 4 is the large disparity among reanalysis in terms of trends, being for example most CFSR-driven products negatively correlated with ERA5, which means that these datasets are exhibiting opposite patterns of trends. Spatial correlation among products is in general medium to low for most products with a correlation lower than 0.6 (the correlation is between 0 and 0.2 for d4PDF-WaveHs members). These discrepancies have been noted in previous studies and are arguably related to the changes in data assimilation over time in reanalysis products^{19,35,41}.

Overall, the d4PDF-WaveHs performance is reasonable given the uncertainty among modern wave data products. The mean climatology is comparable to that obtained from modern reanalysis while extremes tend to be underestimated (especially in tropics), as similarly obtained for the other historical simulations used here for comparison that do not consider data assimilation (CMIP5-COWCLIP, including CMIP5-ECCC(s)). It is particularly challenging to assess the performance of trends given the existing discrepancies among reanalysis. Further data validation is advisable to assess other wave features.

Usage Notes

The data can be used with a wide range of postprocessing software, such as Ferret or NCL, and several packages from Python, R or MATLAB among others.

Code availability

Sub-daily data generation code. The technical details of the statistical model used to generate the 6-hourly H_s from SLP predictors are included in the corresponding reference paper¹⁶ which allow for the reproducibility of the presented dataset. Additionally, the corresponding Fortran and R codes are publicly available in the Government of Canada Open Data Portal, together with the d4PDF-WaveHs dataset (DOI <https://doi.org/10.18164/d68361d0-8141-48b9-a25e-a9bc98d71438>).

Computation of statistics/indices: *getStat.f*, *getHsEx.f*. To be consistent with COWCLIP2.0⁸, the H_s statistics and indices were computed with *getStat.f* and *getHsEx.f*, after a slight modification to account for missing data (see Methods Section). The original Fortran code²⁵ was developed as part of the COWCLIP community framework and can be accessed via the COWCLIP website (<https://cowclip.org/data-access>). The code can be compiled with a Fortran compiler, with netCDF4 and HDF5 libraries. Additional attribute information to account for CF conventions and ACCD standards, was added to this Fortran code output with standard NetCDF operators (NCOs) for file manipulation, such as the “ncatted” command.

Received: 13 December 2022; Accepted: 8 March 2023;

Published online: 06 June 2023

References

- Cavaleri, L., Fox-Kemper, B. & Hemer, M. Wind waves in the coupled climate system. *Bulleting of the American Meteorological Society* **93**, 1651–1661, <https://doi.org/10.1175/BAMS-D-11-00170.1> (2012).
- Gudmestad, O. T. Modelling of waves for the design of offshore structures. *Journal of Marine Science and Engineering* **8**, 293, <https://doi.org/10.3390/jmse8040293> (2020).
- Stive, M. J. *et al.* Variability of shore and shoreline evolution. *Coastal Engineering* **47**, 211–235, [https://doi.org/10.1016/S0378-3839\(02\)00126-6](https://doi.org/10.1016/S0378-3839(02)00126-6) (2002).
- Grifoll, M., Martínez de Osés, F. & Castells, M. Potential economic benefits of using a weather ship routing system at short sea shipping. *WMO Journal of Maritime Affairs* **17**, 195–211, <https://doi.org/10.1007/s13437-018-0143-6> (2018).
- Reguero, B., Losada, I. & Méndez, F. A recent increase in global wave power as a consequence of oceanic warming. *Nature Communications* **10**, 205, <https://doi.org/10.1038/s41467-018-08066-0> (2019).
- Griggs, G. & Reguero, B. Coastal adaptation to climate change and sea-level rise. *Water* **13**, 2151, <https://doi.org/10.3390/w13162151> (2021).
- IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T. F. *et al.* (eds.)] (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp. 2013).
- Morim, J. *et al.* A global ensemble of ocean wave climate projections from cimp5-driven models. *Nature Scientific Data* **7**, <https://doi.org/10.1038/s41597-020-0446-2> (2020).
- Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E. & Jahn, A. Influence of internal variability on arctic sea-ice trends. *Nature Climate Change* **5**, 86–89 (2015).
- Milinski, S., Maher, N. & Olonscheck, D. How large does a large ensemble need to be? *Earth Systems Dynamics* **11**, 885–901, <https://doi.org/10.5194/esd-11-885-2020> (2020).
- Morim, J. *et al.* A global ensemble of ocean wave climate statistics from contemporary wave reanalysis and hindcasts. *Nature Scientific Data* **9**, 358 (2022).
- Maher, N., Milinski, S. & Ludwig, R. Large ensemble climate model simulations: introduction, overview, and future prospects for utilizing multiple types of large ensemble. *Earth System Dynamics* **12**, 401–418, <https://doi.org/10.5194/esd-12-401-2021> (2021).
- Kirchmeier-Young, M., Zhang, X. & Wan, H. Climate change attribution with large ensembles. In EGU21-3404 (ed.) *EGU General Assembly*, 19–30, <https://doi.org/10.5194/egusphere-egu21-3404,2021> (online, 2021).
- Santos, V. M. *et al.* Statistical modelling and climate variability of compound surge and precipitation events in a managed water system: a case study in the netherlands. *Hydrology and Earth System Sciences* **25**, 3595–3615, <https://doi.org/10.5194/hess-25-3595-2021> (2021).
- Ishii, M. & Mori, N. d4pdf: large-ensemble and high-resolution climate simulations for global warming risk assessment. *Progress in Earth and Planetary Science* **7**, <https://doi.org/10.1186/s40645-020-00367-7> (2020).
- Wang, X. L., Feng, Y. & Swail, V. R. North atlantic wave height trends as reconstructed from the twentieth century reanalysis. *Geophysical Research Letters* **39**, L18705, <https://doi.org/10.1029/2012GL053381> (2012).
- Wang, X. L., Feng, Y. & Swail, V. R. Changes in global ocean wave heights as projected using multimodel cimp5 simulations. *Geophysical Research Letters* **41**, 1026–1034, <https://doi.org/10.1002/2013GL058650> (2014).
- Mizuta, R. *et al.* Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bulletin of the American Meteorological Society* **98**, 1383–1398, <https://doi.org/10.1175/BAMS-D-16-0099.1> (2017).
- Casas-Prat, M. *et al.* Effects of internal climate variability on historical ocean wave height trend assessment. *Frontiers in Marine Science* **9**, <https://doi.org/10.3389/fmars.2022.847017> (2022).
- vonStorch, H. & Zwiers, F. *Statistical analysis in climate resaerch* (Cambridge University Press, Cambridge, United Kingdom, 1999).
- Box, G. & Cox, D. An analysis of transformation (with discussion). *J. R. Stat. Soc. Ser. B* **26**, 211–246 (1964).
- Dee, D. *et al.* The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**(656), 553–597, <https://doi.org/10.1002/qj.828> (2011).
- Herbach, H. *et al.* The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049, <https://doi.org/10.1001/qj.3803> (2020).
- Mizuta, R. *et al.* Climate simulations using mri-agcm3.2 with 20-km grid. *Journal of the Meteorological Society of Japan. Ser. II* **90**, 233–258, <https://doi.org/10.2151/jmsj.2012-A12> (2012).
- Wang, X., Feng, Y., Chan, R. & Mentaschi, L. Quick guide for using the getstat, getstatdir, and getsex fortran codes. Tech. Rep. (2017).
- Casas-Prat, M. *et al.* d4PDF-WaveHs: the first SMILE-based ensemble of global historical wave height. *Government of Canada Open Data Portal*, <https://doi.org/10.18164/d68361d0-8141-48b9-a25e-a9bc98d71438> (2023).
- Wang, X. L., Feng, Y. & Swail, V. R. Climate change signal and uncertainty in cimp5-based projections of global ocean surface wave heights. *Journal of Geophysical Research: Oceans* **120**, 3859–3871, <https://doi.org/10.1002/2015JC010699> (2015).
- Morim, J. *et al.* Robustness and uncertainties in global multivariate wind-wave climate projections. *Nature Climate Change* **9**, 711–718, <https://doi.org/10.1038/s41558-019-0542-5> (2019).
- Morim, J. *et al.* Global-scale changes to extreme ocean wave events due to anthropogenic warming. *Environmental Research Letters* **16**, 074056, <https://doi.org/10.1088/1748-9326/AC1013> (2021).
- Reguero, B., Menéndez, M., Méndez, F., Mínguez, R. & Losada, I. A global ocean wave (gow) calibrated reanalysis from 1984 onwards. *Coastal Engineering* **65**, 38–55, <https://doi.org/10.1016/j.coastaleng.2012.03.003> (2012).
- Hemer, M., Katzfey, J. & Trenham, C. Global dynamical projections of surface ocean wave climate for a future high greenhouse gas emission scenario. *Ocean Modelling* **70**, 221–245, <https://doi.org/10.1016/j.ocemod.2012.09.008> (2013).
- Smith, G. E. Global wave hindcast with australian and pacific island focus: fromk past to present. *Geoscience Data Journal* **n/a**, <https://doi.org/10.1002/gdj3.104> (2020).
- Perez, J., Menendez, M. & Losada, I. Gow2: A global wave hindcast for coastal applications. *Coastal Engineering* **124**, 1–11, <https://doi.org/10.1016/j.coastaleng.2017.03.005> (2017).
- Metaschi, L., Vousdoukas, M., Voukoulalas, E., Dosio, A. & Feyen, L. Global changes of extreme coastal wave energy fluxes triggered by intensified teleconnection patterns. *Geophysical Research Letters* **44**, 2416–2426, <https://doi.org/10.1002/2016GL072488> (2017).
- Stopa, J. E., Ardhuin, F., Stutzmann, E. & Lecocq, T. Sea state trends and variability: consistency between models, altimeters, buoys and seismic data (1979–2016). *Journal of Geophysical Research: Oceans* **124**, 2923–3940, <https://doi.org/10.1029/2018JC014607> (2019).
- Bricheno, L. & Wolf, J. Future wave conditions of europe, in response to high-end climate change scenarios. *Journal of Geophysical Research: Oceans* **123**, 8762–8791, <https://doi.org/10.1029/2018JC013866> (2018).
- Bidlot, J.-R., Lemos, G. & Semedo, A. Era5 reanalysis and era5 based ocean wave hindcast. *2nd International Workshop on Waves, Storm Surges, and Coastal Hazards - 16th International Workshop on Wave Hindcasting and Forecasting (2019)* (2019).
- Timmermans, B. W., Gommenginger, C. P., Dodet, G. & Bidlot, J.-R. Global Wave Height Trends and Variability from New Multimission Satellite Altimeter Products, Reanalyses, and Wave Buoys. *Geophysical Research Letters* **47**, e2019GL086880, <https://doi.org/10.1029/2019GL086880> (2020).
- Mori, N., Shimura, T., Kamahori, H. & Chawla, A. Historical wave climate hindcasts based on jra-55. *Coastal Dynamics* (2017).

40. Shimura, T., Mori, N. & Hemer, M. Variability and future decreases in winter wave heights in the western north pacific. *Geophysical Research Letters* **43**, 2716–2722, <https://doi.org/10.1002/2016GL067924> (2016).
41. Sharmar, V. D., Markina, M. Y. & Gulev, S. K. Global ocean wind-wave model hindcasts forced by different reanalyzes: a comparative assessment. *Journal of Geophysical Research Oceans* **126**, e2020JC016710, <https://doi.org/10.1029/2020JC016710> (2021).
42. Shimura, T. & Mori, N. High-resolution wave climate hindcast around japan and its spectral representation. *Coastal Engineering* **151**, 1–9, <https://doi.org/10.1016/j.coastaleng.2019.04.013> (2019).
43. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *Journal of geophysical Research Atmospheres* **106**, 7183–7192, <https://doi.org/10.1029/2000JD900719> (2001).

Acknowledgements

This study utilized the d4PDF, which was produced using the Earth Simulator as Strategic Project with Special Support of JAMSTEC in cooperation with Integrated Research Program for Advancing Climate Models (TOUGOU), Social Implementation Program on Climate Change Adaptation Technology (SI-CAT), and Data Integration and Analysis System (DIAS) funded by the Ministry of Education, Culture, Sport, Science, and Technology (MEXT), Japan.

Author contributions

M.C.P. wrote the manuscript. M.C.P and X.W. co-designed/lead the study. Y.F. contributed to the statistical data analysis. X.W. conceived and produced the original sub-daily d4PDF-WaveHs data with the Disaster Prevention Research Institute (DPRI) of Kyoto University providing funding support for her two-month sabbatical at DPRI hosted by N.M. Y.F., R.C., N.M. and T.S. contributed to the production of the d4PDF-WaveHs dataset. N.M. and T.S. provided the d4PDF atmospheric data and WW3 wave simulations used for validation. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02058-6>.

Correspondence and requests for materials should be addressed to M.C.-P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© Environment and Climate Change Canada 2023